

Learning Flexible Hyperspectral Features

Abdulrahman Galal¹, Hesahm Hasan²

¹Science and Technology Group, Arab Administrative Development Organization, Cairo, Egypt

²Computer Science Department, Faculty of Computers and Information, Cairo University, Cairo, Egypt
abdo1234@msn.com

Abstract- Most researches in the hyperspectral field adjust feature extraction techniques as major dimensionality reduction tools. This is due to the high dimensionality problems. Feature extraction techniques are not limited to such purpose but extended to handle the changing spectral responses. To achieve that, these techniques transform the spectral response into a new domain where features are arranged according to specific criterion. Each technique extracts unique features that are totally different to that others extract. Besides, each technique has advantages and disadvantages regarding handling the highly mixed datasets and the small training sample size. Therefore, utilizing a technique than another may lead to significant information loss. To overcome this problem and derive flexible features, the proposed approach combines the resulting features of each extraction technique in one feature vector and employs a Support Vector Machine (SVM) to classify it. The feature vector consolidates the benefits of each individual technique and neutralizes their disadvantages. Minimum Noise Fraction (MNF), Principle Component Analysis (PCA) and Independent Component Analysis (ICA) have been used in the proposed approach. Experimental results show that the proposed approach overcomes the traditional feature extraction techniques.

Keywords- *Hyperspectral Classification; Feature Extraction; Data Mining; Learning; Support Vector Machine*

I. INTRODUCTION

The high dimensionality problems of hyperspectral images led to utilizing feature extraction techniques as major dimensionality reduction tools. To realize how severe these problems are, two images, hyperspectral and multispectral, are considered. The data volume of the hyperspectral image is 50 times larger than the multispectral one with the same spatial coverage. As a result, the hyperspectral image size is gigantic compared to few megabytes for multispectral image. Instead of treating each pixel in multispectral images as a 7 dimensional vector, we have to deal with a vector of 242 dimensions for each hyperspectral pixel.

However, feature extraction techniques are not limited to dimensionality reduction purpose but extended to handle the changing nature of material spectral responses and the highly mixed classes (Chang, 2000). These techniques transform the material spectral responses to a new domain where features are arranged according to specific criterion such as data variance. Each feature extraction technique has unique transformation properties. For example, Principal component analysis (PCA) transforms the data according to variance (Richards, 1999; Rodarmel & Shan, 2002). Minimum noise fraction (MNF) transforms the data according to signal-to-noise ratio (SNR) (Weizman & Goldberger, 2007). Independent component analysis (ICA) transforms the data into maximally independent components (Bayliss et al., 1997). Each technique has advantages and disadvantages regarding handling the highly mixed datasets and the small training sample size. Hence, it is hard to decide which technique is appropriate to handle specific situation. The proposed approach resolved this issue by consolidating their advantages and neutralizing their

disadvantages. The proposed approach consists of two stages. In the first stage, MNF is applied on the training dataset to extract the highest quality bands. In the second stage, PCA and ICA bands are applied on the MNF bands of the first stage. For each pixel in the training dataset assigned to specific class, the corresponding PCA and ICA values of this pixel are combined in one feature vector. The feature vector is classified using SVM with the corresponding class label. The proposed approach has been tested against individual feature extraction techniques on a benchmark dataset.

The paper is organized as follows: Section II presents an overview of the investigated features of extraction techniques; Section III describes the proposed approach; Section IV discusses the performance of the proposed approach and the competitive techniques; and finally the conclusions are given in Section V.

II. SPECTRAL TRANSFORMATIONS

A. Principal Component Analysis (PCA)

PCA transforms the highly correlated bands into a new domain where the transformed bands are uncorrelated and arranged according to their variance. The output of PCA is called PC bands. The first PC band contains the largest percentage of data variance and the second PC band contains the second largest data variance, and so on. The last PC bands appear noisy because they contain very little variance. PCA segregates noise components and reduces the dimensionality of hyperspectral data.

PCA is done through finding a new set of orthogonal axes that have their origin at the data mean and that are rotated to maximize the data variance. PCA requires the computation of the eigenvectors and eigenvalues of the covariance matrix. The covariance matrix Σ is defined as:

$$\Sigma = \sum_{i=1}^N (\bar{x}_i - \bar{m})(\bar{x}_i - \bar{m})^T \quad (1)$$

Where \bar{x}_i is the i^{th} spectral signature, \bar{m} denotes the mean spectral signature, and N is the number of image pixels. In order to find the new orthogonal axes of the PCA space, eigen decomposition of the covariance matrix Σ is performed and is defined as:

$$\Sigma \bar{a}_k = \lambda_k \bar{a}_k \quad (2)$$

Where λ_k is the K^{th} eigenvalue, \bar{a}_k denotes the corresponding eigenvector and k varies from 1 to the number of hyperspectral images bands. The eigenvectors form the axes of the PCA space, and they are orthogonal to each other. The eigenvalues denote the amount of variance of the corresponding eigenvectors and they are arranged in decreasing order of the variance. Consequently, the first PC bands are retained as they contain a significant level of information. PCA transformation matrix, A , is formed by choosing the eigenvectors corresponding to the largest eigenvalues and A is defined as:

$$A = [\bar{a}_1 | \bar{a}_2 | \dots | \bar{a}_J] \quad (3)$$

Where $\bar{a}_1 | \bar{a}_2 | \dots | \bar{a}_J$ are the eigenvectors associated with the J largest eigenvalues obtained from the eigen decomposition of the covariance matrix Σ . The data projected onto the corresponding eigenvectors form the reduced uncorrelated features that are used for further classification processes.

To realize how PCA works, it is assumed that the significant feature of a target is its high reflectance value in red and near-infrared (NIR) regions while the non-target is discriminated by its low reflectance in either red or NIR region as shown in Fig. 1.

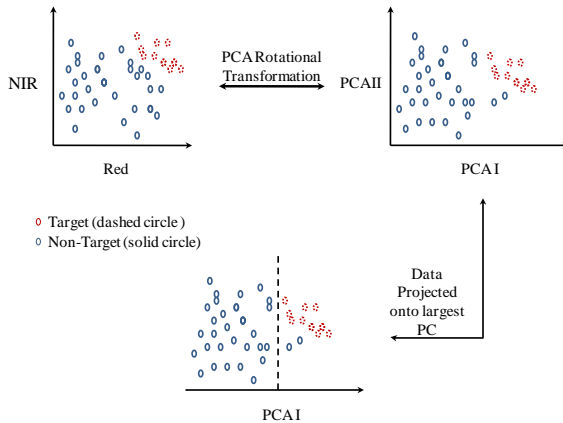


Fig. 1 PCA transformation example

This figure depicts how the information in the red and NIR bands have been compressed to form a single dimension, which is the largest principal component (PCA I). PCA I is calculated using the eigen decomposition of the covariance matrix estimated from the data-distribution. By measuring the data projected onto this new dimension, objects having high reflectance in the red and NIR bands (targets) can be easily discriminated from objects having low reflectance values either in the red or NIR band. The amount of information that is retained by PCA I dimension can be determined from the eigenvalues. PCA results in establishing the lower-dimensional projection that maximizes the variance present in the data.

B. Minimum Noise Fraction Transform (MNF)

Real sensor noise is not isotropic. Isotropic noise means that the random noise radiation reaches a location from all directions with equal intensity. Signal-to-noise ratio (SNR) is a measure of image quality. SNR compares the level of a desired signal to the level of background noise. MNF takes into account sensor noise and orders the images in terms of SNR. In contrast, PCA considers only the variances of each PC component rather than sensor noise. PCA assumes that noise is isotropic (Rodarmel, 2002). MNF transform consists of two cascaded principal component transformations. The first transformation is based on the estimated noise covariance matrix. It transforms data in which the noise has unit variance and no band-to-band correlations. The second transformation is a standard principal components transformation of the noise-whitened data. The resulting eigenvalues and the corresponding images from the transformations are called Eigen Images. Eigen Images associated with large eigenvalues contain useful information while eigenvalues close to one indicate noise dominated data. The first transform is derived

from the covariance matrix for the sensor noise, and is designed to de-correlate and whiten the data with respect to the noise. Therefore, it is required to estimate the signal and noise covariance matrices. The main difficulty with this process is obtaining a proper estimate of the sensor noise. Many noise estimation methods have been suggested: (1) Simple Differencing: noise is estimated by differencing the image between adjacent pixels; (2) Simultaneous Auto Regressive: noise is estimated by the residual in a SAR model based on W, NW, E and NE pixels; (3) differencing with the local mean and (4) differencing with the local median.

C. Independent Component Analysis (ICA)

ICA assumes that each band is a linear mixture of independent hidden components and proceeds to recover the original factors or independent features through a linear unmixing operation. Let x_t and s_t denote the linear mixtures and original source signals respectively; the aim of the ICA is to estimate s_t by:

$$s_t = Ux_t \quad (4)$$

where U is unmixing matrix. For estimating s_t , ICA assumes s_t components are independent statistically and all of them with possible exception of one component must be non-Gaussian. Hence, it needs higher order information of the original inputs rather than the second-order information of the sample covariance as used in PCA. Fig. 2 shows how PCA and ICA project and cluster the data.

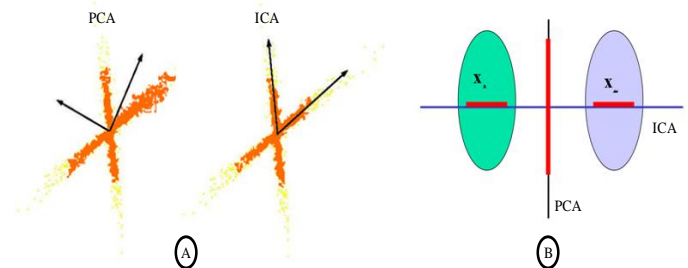


Fig. 2a) PCA (Orthogonal Coordinates) and ICA (Non-Orthogonal Coordinates); b) Clusters created by PCA and ICA

Many algorithms have been developed for performing ICA. One of the best methods is the fixed-point-Fast ICA algorithm (Jin & Harold, 2009). Fast ICA algorithm is based on minimization of mutual information to estimate s_t . Minimization of mutual information is a natural measure of the independence between random variables and corresponds to maximization of entropy which is approximated by:

$$J_G(u_i) = [E\{G(u_i^T x_t)\} - E\{G(v)\}]^2 \quad (5)$$

Where u_i is an m -dimensional vector, comprising one of the rows of the matrix U , v is a standardized Gaussian variable and G is a non-quadratic function. Maximizing $JG(u_i)$ leads to estimating u_i by:

$$u_i^+ = E\{x_t g(u_i^T x_t)\} - E\{g'(u_i x_t)\} u_i \quad (6)$$

$$u_i^* = \frac{u_i^+}{\|u_i^+\|} \quad (7)$$

Where u_i^* is a new estimated of u_i and g , g' are first and second derivative of G . At every iteration, the vectors $u_i^* x$ are decorrelated using a symmetric decorrelation of the matrix U :

$$U = (UU^T)^{-1/2} U \quad (8)$$

With U matrix is composed of $(u_1, u_2, \dots, u_n)^T$ of vector u_i and $(UU^T)^{-1/2}$ is obtained by the eigenvalue decomposition of U .

III. THE PROPOSED APPROACH

The proposed approach consolidates the advantages of PCA, MNF and ICA and neutralizes their disadvantages. The proposed approach consists of two cascaded stages. In the first stage, MNF is applied to the dataset. The results are MNF bands ordered by SNR. The first 10 MNF bands are selected as the basis for the next stage. In the second stage, PCA and ICA are applied on the selected MNF bands from stage 1. The results are 20 bands, 10 for PCA and 10 for ICA. Finally, for each training sample or pixel, a vector composed of 20 values (10 values for the corresponding PCA band pixel and 10 values for the corresponding ICA band pixel) is formed. The resulting vector is attached with class label and classified by SVM. Fig. 3 depicts the proposed approach. experimental evaluation

A. Dataset

The dataset represents an Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) image with size of 145×145 pixel vectors taken from an area of mixed agriculture and forestry in Northwestern Indiana, USA. This dataset has been chosen because it has been studied extensively in hyperspectral classification field. It should be noted that, at the

time of data acquisition, the SNR was considerably lower than current AVIRIS standards. The data was recorded with 220 bands with water absorption bands. Bands 104-108 and 150-162 are removed, leaving only 202 bands. (The URL of the dataset is <ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.tif.zip>)

The scene is categorized into 17 classes as shown in Fig. 4. Many methods claiming to work well on classification were not able to break down this image scene. This is because pixels in the same class are highly mixed that any spectral similarity measure may consider pixels in different classes be considered to belong to the same class.

B. Experimental Methodology

ENVI 4.5 program has been used for generating PCA, MNF and ICA bands. The top 10 bands of each technique were selected. PCA bands were arranged according to data variance. MNF bands were arranged according to SNR values. ICA bands were arranged according to the D spatial coherence. For classification, LIBSVM (Chih & Chih-Jen, 2011) has been used as SVM tool. LIBSVM parameters were: (Multi-Class-Type: One-Against-One) (Kernel Type: Radial Basis Function) (Penalty Parameter: 1000.00). These parameters have been derived from Watanachaturaporn et al. (2004) and Camps-Valls et al. (2004) researches that suggest the best parameters of SVM applied on the same test dataset.

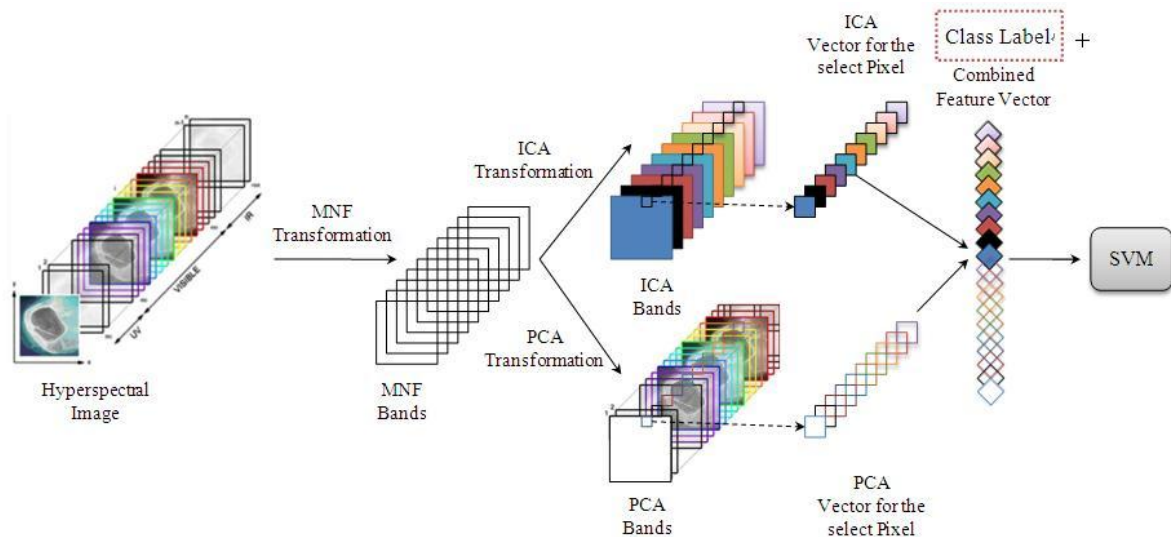


Fig. 3 The proposed approach

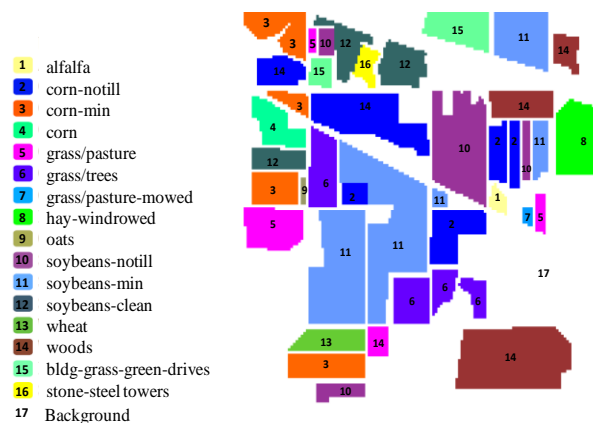


Fig. 4 The test area and the ground truth classes

C. Results

According to the analysis conducted by Wu and Chang (2009) on the test dataset, the spectral signatures of classes (2, 3, 4, 7, 10 and 12) are so close to each other, and the same condition for classes (1, 8, and 11). For classes (5, 14, and 15), they have less similar signatures. For classes (6, 13 and 16), their signatures are dissimilar. Classes 5 and 11 are highly mixed. The classification rates of PCA, MNF, ICA and the proposed approach are listed in Table I.

TABLE I CLASSIFICATION RATES OF PCA, MNF, ICA AND THE PROPOSED APPROACH

Class	Training	Test	PCA	MNF	ICA	Proposed Approach
1	35	19	15.02	16.10	25.13	75.05
2	742	692	91.91	80.81	93.5	95.02
3	442	392	81.07	70.12	85.20	90.15
4	180	54	55.48	46.33	58.02	60.15
5	260	237	96.62	90.35	93.02	98.62
6	389	358	99.44	99.31	99.40	99.5
7	20	6	8.15	4.23	15.05	71.75
8	236	253	99.60	100	100	100
9	15	5	10.02	7.02	18.01	61.05
10	487	481	68.19	70.02	75.05	79.02
11	1245	1223	94.94	91.92	95.01	94.94
12	305	309	96.12	72.20	97.03	97.9
13	150	62	35.48	30.33	41.19	65.05
14	651	643	99.53	99.50	99.81	99.41
15	200	180	44.13	40.50	55.06	56.02
16	55	40	33.56	32.35	31.06	65.02
17	6659	4000	70.03	60.02	79.02	91.38
Avg. Classification Rate			78.32	70.71	83.75	91.30

The discussions of the results are as follows:

For PCA and MNF: MNF performance was the worst. PCA performed better than MNF. The explanation of their low performance was due to the following reasons:

1. The samples of classes (1, 4, 7, 9, 13, and 15) are relatively small and not sufficient to constitute reliable statistics. In this case, these classes are not captured by the second-order statistics-based PCA in its PCs. As a result, PCA and MNF are not able to correctly preserve the information of interest as shown in Fig5. (a), (b).

2. Ignoring the lower-order PC components of the class (3 and 10) resulted in losing some of the discriminatory information.

3. The noise in the dataset is so high that it was transparent by comparing the eigenvalues of both bands generated by PCA and NFA as shown in Table II. The huge gap in eigenvalues between the 1st and 2nd PCA confirms also the presence of high noise. PCA band 8 in Fig. 5(a) was totally distorted.

For ICA: ICA performance was poor in small size classes as PCA and MNF but it achieved a better average classification rate than PCA and MNF. This is due to the following reasons:

1. ICA was able to classify samples where the signal of interest is relatively weak compared to other signals in the data.

This happened in classes 1, 9 and 13 as shown in Fig. 5(c) ICA bands 1, 5 and 7 respectively.

2. ICA bands are not sorted in the same way of PCA and MNF. Instead, ENVI sorts IC bands according to 2D spatial coherence. Accordingly, the top ICA bands combined two advantages: the high data variance and sources separation information.

3. ICA assumes that each band is a linear mixture of independent hidden components and proceeds to recover the original factors or independent features through a linear unmixing operation. Hence, ICA achieved better in classes (1, 4, 7, 9, 13, and 15) compared to PCA and MNF.

4. ICA performance considered good across all the classes especially the highly mixed classes such as 5 and 11.

TABLE II PCA AND MNF STATISTICS

Bands	PCA		MNF	
	Stdev	Eigenvalue	Stdev	Eigenvalue
Band 1	5038.16	25383111.68	6.821972	46.53930
Band 2	2937.63	8629687.529	4.397703	19.33979
Band 3	778.735	606428.6292	3.848766	14.81299
Band 4	358.521	128537.9653	3.570652	12.74955
Band 5	263.203	69275.82620	3.452201	11.91769
Band 6	230.806	53271.49487	2.993792	8.962792
Band 7	161.196	25984.19208	2.849855	8.121672
Band 8	116.696	13617.97909	2.552359	6.514535
Band 9	112.565	12670.98120	2.415808	5.836127
Band 10	91.3024	8336.142599	2.189476	4.793805

For the proposed approach: The proposed approach achieved the best results among the rest approaches. This is due to the following reasons:

1. Applying MNF as a prerequisites step before calculating PCA and ICA bands enable handling the small size class very well compared. This means the performance of PCA and ICA is dramatically improved if they are applied on the highest quality bands or MNF bands. By comparing the results of PCA results applied on the raw data and PCA results applied on the MNF bands, it is clear that the proposed approach avoided the high noise band as shown in Table III.

TABLE III PCA STATISTICS AFTER APPLYING MNF

Bands	Stdev	Eigenvalue
Band 1	6.821972	46.539306
Band 2	4.397703	19.339791
Band 3	3.848766	14.812997
Band 4	3.570652	12.749556
Band 5	3.452201	11.917691
Band 6	2.993792	8.962792
Band 7	2.849855	8.121672
Band 8	2.552359	6.514535
Band 9	2.415808	5.836127
Band 10	2.189476	4.793805

2. It combined the strengths of ICA to unmix the highly mixed, richness of PCA data variance and MNF to remove highly noise bands. Such combination is considered flexible

feature inputs that helped SVM to learn how to discriminate the classes from each other. This is highly transparent in the sample of bands shown in Fig. 3(d) band 1 to 8.

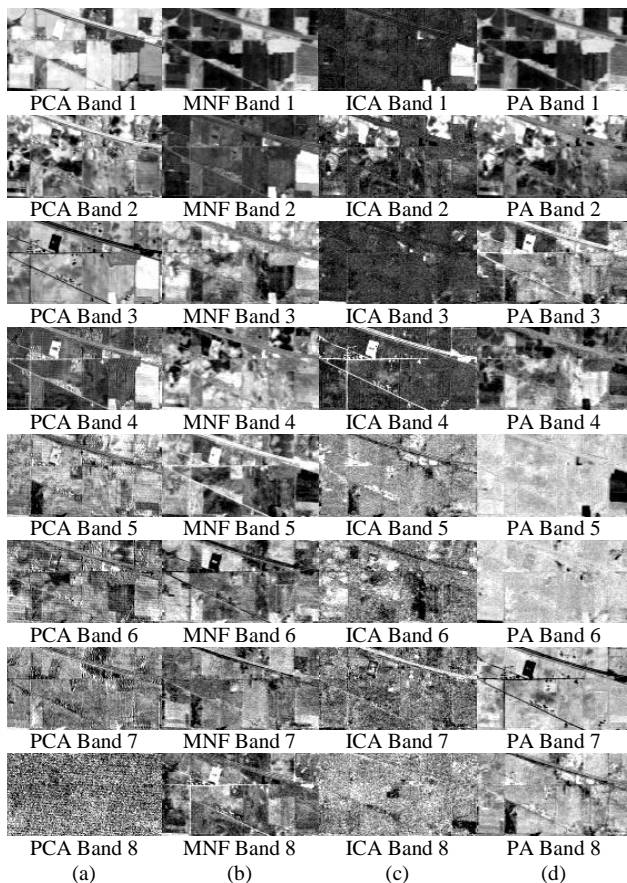


Fig. 5 Samples of a) PCA bands b) MNF bands c) ICA bands d) Proposed Approach (PA) sample bands

IV. CONCLUSIONS

Each feature extraction technique extracts unique features that are totally different with what others extract. Depending on a technique than another may lead to significant information loss. The proposed approach combines the resulting features of each extraction technique in one feature vector and employs a Support Vector Machine (SVM) to classify it. The flexible feature vector helps SVM to learn well how to discriminate the classes from each other. According to

experiment results, PCA and MNF performance was the worst compared to ICA and the proposed approach. The proposed approach achieved the best among the rest approaches. It combined the strengths of ICA to unmix the highly mixed, richness of PCA data variance and MNF to remove highly noise bands. Besides, It performed very well on small training sample size and the highly mixed classed. This is due to applying MNF as a prerequisites step before calculating PCA and ICA bands. The performance of PCA and ICA is dramatically improved when they are applied on the highest quality bands or MNF bands.

REFERENCES

- [1] G., Camps-Valls, L., Gomez-Chova, J., Calpe-Maravilla, J., Mart'ín-Guerrero, J., E., Soria-Olivas, L., Alonso-Chorda, J., Moreno, J., "Robust support vector technique for hyperspectral data classification and knowledge discovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42(7), pp. 1530-1542, 2004.
- [2] J., Bayliss, J., Gualtieri, J., Cromp, R., "Analyzing hyperspectral data with independent component analysis," *SPIE*, vol. 3240, pp. 133-143, 1997.
- [3] C., Chang, "An Information Theoretic-based Approach to Spectral Variability, Similarity and Discriminability for Hyperspectral Image Analysis," *IEEE Trans. Information Theory*, vol. 46(5), pp. 1927-1932, 2000.
- [4] C., Chih-Chung, L., Chih-Jen, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology Journal*, vol. 2(3), pp. 1-27, 2011.
- [5] X., Jin, P., Scott, C., Harold, C., "A comparative study of target detection algorithms for hyperspectral imagery," *SPIE*, vol. 73341W, 2009.
- [6] C., Mingmin, K., Qian, A., Jon, F., Rui, "Ensemble Classification Algorithm for Hyperspectral Remote Sensing Data," *IEEE Geoscience and Remote Sensing Letters*, 3, pp. 762-766, 2009.
- [7] J. Richards, "Remote Sensing Digital Image Analysis: An Introduction," Springer-Verlag, pp. 240, 1999.
- [8] C., Rodarmel, J., Shan, "Principal Component Analysis for Hyperspectral Image Classification," *Journal of the American Congress on Surveying and Mapping*, 2, pp.115-122., 2002.
- [9] P. Watanachaturaporn, P., Varshney, M., Arora, "Evaluation of factors affecting support vector machines for hyperspectral classification," *American Society for Photogrammetry & Remote Sensing (ASPRS) 2004 Annual Conference*, 2004.
- [10] L., Weizman, J., Goldberger, "A Classification Based Linear-Projection of Labeled Hyperspectral Data," *27th International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3202-3205, 2007.
- [11] C., Wu, C., Chang, "Soft Decision-made Hyperspectral Measures for Target Discrimination and Classification," *SPIE* 7457, 74570S, 2009.